The Role of Corpus Pattern Analysis in Machine Translation Evaluation

Hanna Béchara
University of Wolverhampton
hanna.bechara*

Sara Moze
University of Wolverhampton
S.Moze*

Ismail El-Maarouf
University of Wolverhampton
I.el-maarouf*

Constantin Orăsan
University of Wolverhampton
C.Orasan*

Patrick Hanks
University of Wolverhampton
patrick.w.hanks**

Ruslan Mitkov
University of Wolverhampton
R.Mitkov*

*** @WLV.AC.UK** ****@GMAIL.COM**

Keywords: Machine Translation, Corpus Pattern Analysis, pattern dictionary, syntax, semantics, translation evaluation

Abstract

This paper takes a preliminary look at the relation between verb pattern matches in the Pattern Dictionary of English Verbs (PDEV) and translation quality through a qualitative analysis of human-ranked sentences from 5 different machine translation systems. The purpose of the analysis is not only to determine whether verbs in the automatic translations and their immediate contexts match any pre-existing semanto-syntactic pattern in PDEV, but also to establish links between hypothesis sentences and the verbs in the reference translation. It attempts to answer the question of whether or not the semantic and syntactic information captured by Corpus Pattern Analysis (CPA) can indicate whether a sentence is a "good" translation. Two human annotators manually identified the occurrence of patterns in 50 translations and indicated whether these patterns match any identified pattern in the corresponding reference translation. Results indicate that CPA can be used to distinguish between well and ill-formed sentences.

1. INTRODUCTION

In light of recent advancements in machine translation, the question of machine translation evaluation has become more visible. As human evaluation is both costly and time-consuming, developers have relied on a variety of techniques to assess the quality of machine translation output. Many of these automatic evaluation metrics are shallow tools that match parts of the output to a reference translation, and fail to correspond to human judgement (Callison-Burch et al., 2008).

Corpus Pattern Analysis (CPA) is a corpus-driven technique in corpus linguistics and lexicography that associates word meaning with word use by mapping meaning onto specific syntagmatic patterns exhibited by a verb in any type of text. CPA aims at identifying patterns[1] of normal usage ('norms'), including literal and metaphorical uses,

[1] In CPA, a pattern is "*described according to five types of arguments: Subject, Object, Complement, Adverbial, and Indirect Object. Each can be further detailed using, semantic types,*

phrasal verbs and idioms, and exploring the way patterns are creatively exploited ('exploitations') (Hanks, 2013). CPA is currently being used to compile the Pattern Dictionary of English Verbs (PDEV),[2] an online lexical resource that currently covers nearly 1,300 English verbs.

The attractiveness of CPA in machine translation evaluation comes from the fact that CPA provides information on both word senses and the typical syntactic patterns associated with them. For this reason, we argue that employing CPA in the evaluation process will bring significant advantages over the existing methods.

2. MACHINE TRANSLATION EVALUATION MEETS CPA: A PILOT STUDY

2.1. Dataset and Methodology

We extracted all our sample sentences from the dataset prepared for the WMT2013 shared task in Quality Estimation (Bojar et al., 2013). This dataset serves our purposes as it presents German source sentences with 5 distinct machine translations (MT), scored by manual assessors on a scale of 1-5, where 1 is the best and 5 is the worst sentence. For each German sentence, a reference translation (REF) is also provided.

We extracted 10 German sentences and their English translations, leaving us with 50 sentences to annotate in total. Once extracted, we tagged the sentences using TreeTagger (Schmid, 1994) to identify potential patterns in each of the MT sentences. Using Palinka (Orăsan, 2003), we investigated whether or not these patterns match any identified in the reference translation. A pattern could be a full match (occurs in the reference within the right semantic and syntactic context), or a partial match (occurs in the reference, but not within the same context). The annotated data was then analysed using statistical (cf. section 2.2.) and qualitative methods (cf. section 2.3.) to illustrate the advantages of using PDEV as a source of data in MT evaluation.

contextual roles, and lexical sets" (El Maarouf et al., 2014: 1002).

[2] http://www.pdev.org.uk

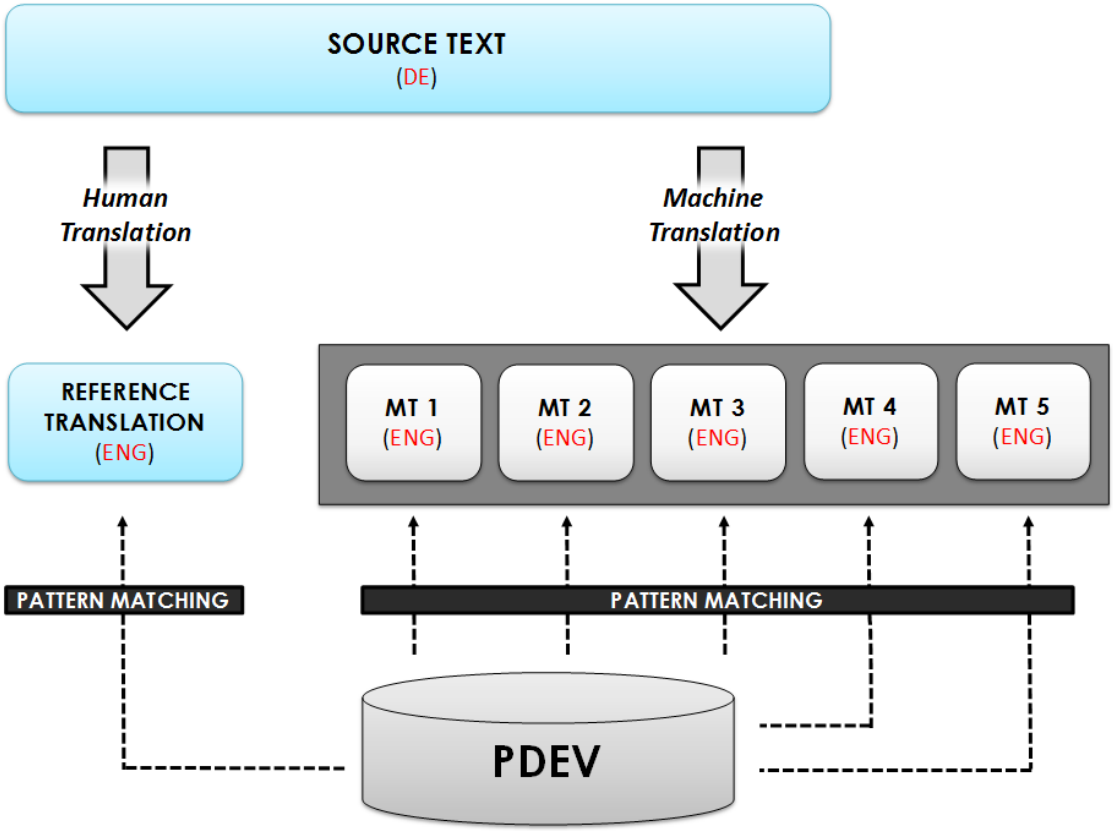

Figure 1 – Methodology

2.2. Results

Our results show a correlation between the ranks and the matched patterns. Machine translated sentences with better rankings (1-3) generally have a higher number of verbs occurring in the same context as the verbs in the given reference translation, as shown in Table 1.

SENTENCE NO.	RANK 1	RANK 2	RANK 3	RANK 4	RANK 5
1	4	4	5	0	0
2	3	3	2	2	N/A
3	4	2	2	2	N/A
4	2	1	3	0	0
5	1	2	N/A	0	0
6	3	3	3	N/A	1
7	2	N/A	N/A	1	1
8	3	N/A	2	2	2
9	0	1	1	1	N/A
10	0	0	N/A	1	N/A
TOTAL	**22**	**16**	**18**	**9**	**4**

Table 1 – Number of verbs in the MT sentence that match verbs in the reference translation (rank 1 – best; rank 5 – worst)

To conclude, our preliminary analysis seems to indicate that the number of matches and their types can indeed be used as an indicator for translation quality.

2.3. Examples of syntactic and semantic mismatch

PDEV can be used to identify ungrammatical sentences and mistakes in the use of collocations by providing a means to compare sentences produced by MT systems to their closest matching patterns of normal usage in PDEV.

A good example of a syntactic mistake in the MT output is exemplified by a sentence featuring the verb *file* (cf. Figure 2). PDEV lists nine separate patterns for the verb; the closest in terms of syntax and semantics is pattern 2, which features a syntactically simple sentence structure (subject-verb-direct object) and describes a situation where an institution acting as a plaintiff or accuser officially presents a document, i.e. lawsuit or an accusation, to a court of law in order to start legal proceedings.[3]

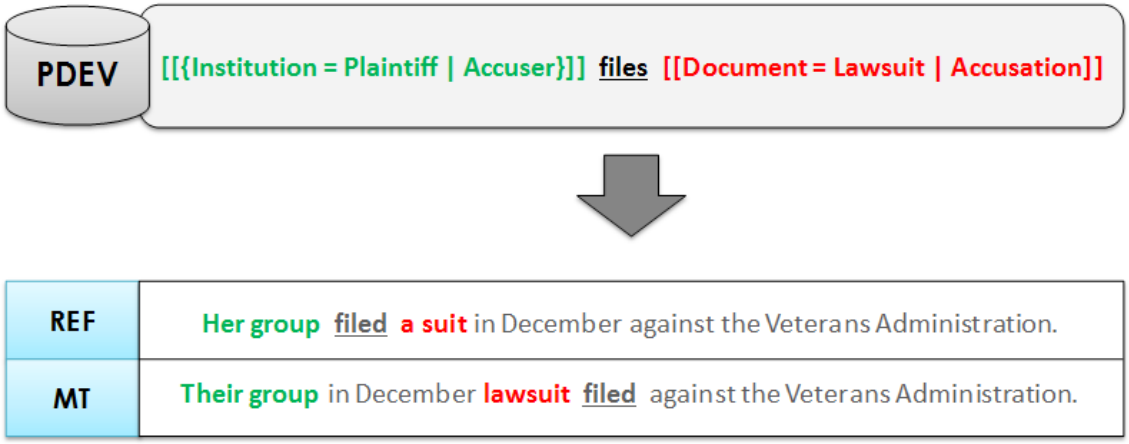

Figure 2 – Identification of syntactic mistakes in MT sentences

Both sentences are lexically and semantically correct, but differ in terms of word order. As the direct object, *lawsuit* should follow the verb and not precede it, which renders the sentence produced by the MT system ungrammatical.

Furthermore, PDEV can be used to pin-point nouns in the MT sentence that do not match the partially matching pattern in terms of semantics. For instance, the near-synonyms *assault* and *abuse* are both used to describe a situation whereby a human is physically and/or sexually attacked by another human (cf. Figure 3). Both patterns feature rather simple syntax and semantics – a sentence structure with a direct object and noun fillers that typically correspond to the semantic type [[Human]]. A *service*, however, is by no means a human being, and sexually abused or assaulted services are hardly something one would expect to encounter in normal English. Semantic mismatches between semantic types and nouns fillers selected by MT systems may

[3] The information between square brackets is as follows: semantic types, e.g. Institution and Document, refer to the semantic category of nouns that typically occur in the specified syntactic slot. They are hierarchically organised in a shallow ontology of semantic types (cf. Jezek and Hanks, 2010). What follows the equals sign are contextual roles, i.e. *"elements of meaning assigned to a word or expression by the context in which it is used"* (Hanks, 2013: 432), e.g. Plaintiff, Accuser, Lawsuit, Accusation.

indicate a semantic shift in the translation – if the meaning of the original is altered significantly, the sentence cannot be considered acceptable.

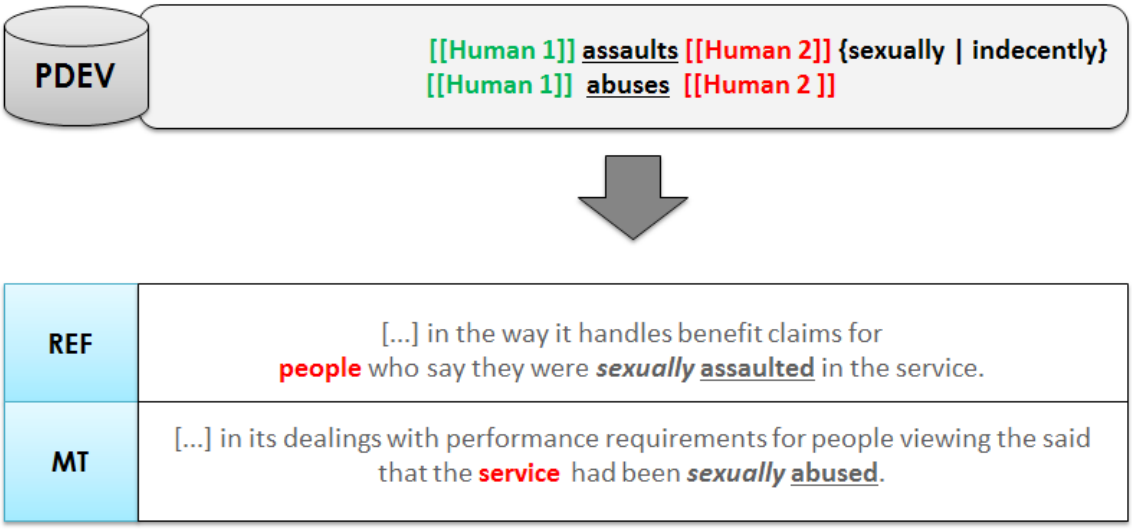

Figure 3 – Identification of semantic mismatches between MT sentences and corresponding patterns in PDEV

A verb's lexical profile cannot be deemed complete without an in-depth look into the verb's full phraseology. Following that view, PDEV provides valuable information on all idioms that are typically associated with a verb by listing them as separate patterns. In some cases, the use of an idiom might not be considered appropriate due to several reasons ranging from stylistic (e.g. register, domain or connotation) to lexicogrammatical (i.e. the semantic and syntactic preferences associated with its use). Consider the following example:

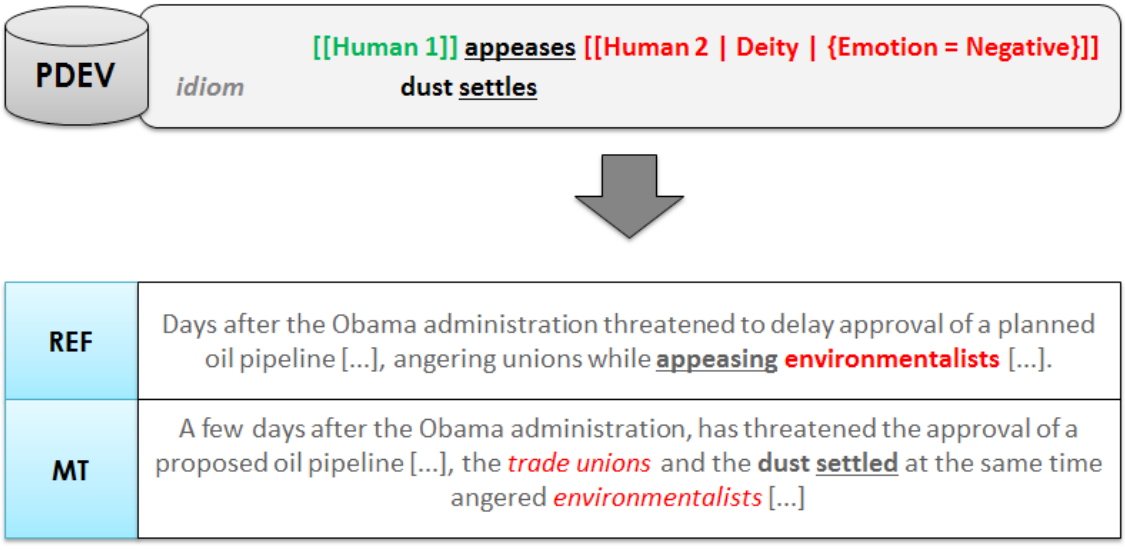

Figure 4 – Incorrect use of an idiom in the MT sentence

The idiom *dust settles* and the verb *appease* both incorporate the notion of 'calm', but are used in different contexts and do not exhibit the same syntactic behaviour. The idiom does not allow for patient-like participants in the direct object slot, which means

that it cannot be combined with the noun *environmentalist* to form a causative construction and thus be used as a good translation equivalent. As a monolingual lexical resource focusing on syntagmatic, rather than paradigmatic, relations, PDEV cannot be expected to provide a commentary on the fine-grained semantic distinctions between potential translation equivalents, the information contained within, however, is more than sufficient to flag the MT sentence in Figure 4 as unacceptable.

3. CONCLUSION

This paper presents a first look into the use of CPA in MT evaluation. In the future, we would like to build upon this work by manually annotating a large number of new sentences and potentially introducing new corpora with the aim of providing much needed variety in terms of text-type, domain and register. The ultimate goal, however, is automation – we believe that developing (semi-) automatic procedures that use PDEV as a source of lexical data to rank sentences would contribute greatly to the development of MT evaluation as a subfield, while integrating PDEV directly into pre-existing MT systems could significantly improve their performance.

Acknowledgements

This work is supported by the People Programme (Marie Curie Actions) of the European Union's Framework Programme (FP7/2007-2013) under REA grant agreement no 317471 and by AHRC grant [DVC, AH/J005940/1, 2012-2015].

References

CALLISON-BURCH, C., FORDYCE, C., KOEHN, P., MONZ, C. AND SCHROEDER, J., 2008. Further Meta-Evaluation of Machine Translation. *Proceedings of the third workshop on statistical machine translation (wmt)*, pp. 70–106.

BOJAR, O., BUCK, C., CALLISON-BURCH, C., FEDERMANN, C., HADDOW, B., KOEHN, P., MONZ, C., POST, M., SORICUT, R. AND SPECIA, L., 2013. Findings of the 2013 Workshop on Statistical Machine Translation. *Proceedings of the Eighth Workshop on Statistical Machine Translation*. Sofia, Bulgaria, pp. 1–42,

EL MAAROUF, I., BAISA, V., BRADBURY, J. AND HANKS, P., 2014. Disambiguating Verbs by Collocation: Corpus Lexicography meets Natural Language Processing. *Proceedings of Euralex 2014*. Reykjavik, pp. 1001–6.

HANKS, P., 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, MA; London: MIT Press.

JEZEK, E. AND HANKS, P., 2010. What lexical sets tell us about conceptual categories. *Lexis: E-journal in English lexicology. 4: Corpus Linguistics and the Lexicon.* Lyon: Université Lumière – Lyon 2, pp. 7–22.

ORĂSAN, C., 2003. PALinkA: a highly customizable tool for discourse annotation. *Proceedings of the 4th SIGdial Workshop on Discourse and Dialog*, pp. 39–43.

The Pattern Dictionary of English Verbs (PDEV). [online] Available at: <http://www.pdev.org.uk> [Accessed 5 March 2015].

SCHMID, H., 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK, pp. pp. 44–49.